



# **INFRAPUNASPEKTRIN ESIKÄSITTELY EMSC- MENETELMÄLLÄ**

Aleksi Leviäkangas

Kandidaatin tutkielma

Hyvinvointitekniikan tutkinto-ohjelma

Lääketieteen tekniikka

Oulun yliopisto

2020

**Leviäkangas Aleks (2020), INFRAPUNASPEKTRIEN ESIKÄSITTELY EMSC-MENETELMÄLLÄ, Lääketieteellinen tiedekunta, Oulun yliopisto, Kandidaatin tutkielma, 33 sivua.**

## **Tiivistelmä**

**Työtarkoitus:** Fourier-muunnos -infrapunamikroskopiolla (FTIR-mikroskopia) voidaan tutkia heterogeenisten näytteiden, kuten kudosten, koostumusta. Valon sironta aiheuttaa kuitenkin huomattavia häiriöitä mitattaviin absorptiospektreihin vaikeuttaen analyysia. Yksi yleisesti esiintyvä sirontatyyppi on Mie-sironta, jota esiintyy varsinkin biologisia näytteitä kuvattaessa. EMSC (extended multiplicative signal correction) on Beer-Lambertin laista johdettu yleisesti käytetty esikäsittelymenetelmä, jolla voidaan poistaa erinäisiä virhetekijöitä, mikäli niitä pystytään jotenkin mallintamaan. Tässä työssä pyritään rakentamaan ketterä EMSC-malli, jolla voidaan poistaa etenkin Mie-sironnan aiheuttamia virheitä absorptiospektreistä.

**Menetelmät:** Muodostettiin Matlabilla EMSC-malli aikaisemman kirjallisuuden perusteella keskittyen erityisesti Mie-sironnan poistamiseen. Testattiin mallia simuloituihin nivelrustospektreihin kuin myös FTIR-mikroskoopilla kuvattuihin nivelrustospektreihin. Verrattiin korjauksen hyvyttä täydellisemmän RMieS-algoritmin (resonant Mie scattering) korjaukseen laskemalla yleisiä tunnuslukuja.

**Tulokset:** Simuloiduilla nivelrustospektreillä muodostettu EMSC-malli vaikutti toimivan moitteettomasti. Oikealla mittausdatalla korjauksessa esiintyi selkeitä puutteita mallin epätäydellisyydestä johtuen.

**Johtopäätökset:** Ketterällä ei-iteratiivisella EMSC-mallilla pystytään kohtuullisen hyvään esikäsittelyyn, mutta parempaan tulokseen päästäisiin tarkemmalla matemaattisella mallinnuksella. Malli sisältää muutamia kriittisiä parametrejä, joiden valintaan tulisi kiinnittää enemmän huomioita.

**Avainsanat:** Infrapunaspektroskopia, FTIR-mikroskopia, nivelrusto, esikäsittely, EMSC, Mie-sironta, RMieS

# Sisällysluettelo

1	Johdanto .....	3
2	Tutkimuksen tausta .....	5
2.1	Infrapunaspektroskopia ja EMSC-malli.....	5
2.2	Mie-sironta .....	7
2.3	Nivelruston kuvantaminen IRS:lla.....	7
3	Tutkimuksen tavoitteet.....	9
4	Materiaalit ja menetelmät.....	10
4.1	Osaspektreillä simuloitu absorptiospektri .....	10
4.2	FTIR-nivelrustospektri .....	16
5	Tulokset.....	19
5.1	Osaspektreillä simuloitu absorptiospektri .....	19
5.1	FTIR nivelrustospektri .....	21
6	Pohdinta .....	29
7	Yhteenveto .....	31
8	Lähteet.....	32

# 1 Johdanto

Infrapunaspektroskopia (IRS) on monikäyttöinen ja mittauskohdetta vahingoittamaton mittausmenetelmä, jolla voidaan määrittää esimerkiksi erinäisten jauheiden ja liuosten kemiallista koostumusta. Yleisimpiä IRS:n tyyppejä ovat Fourier-muunnos infrapunaspektroskopia (FTIR), kvanttikaskadilaserspektroskopia sekä dispersiivinen IR-spektroskopia. IRS perustuu erilaisten kemiallisten sidosten kykyyn absorboida infrapunavaloa (IR) tietyillä aallonpituusalueilla, ts. menetelmä mittaa näytteen molekyylikoostumuksen absorptiospektriä [1], [2]. FTIR-mikroskopiassa laitteistoon on vielä itse spektroskoopin lisäksi kytkettynä valomikroskooppi, jolloin päästään suurempaan spatiaaliseen tarkkuuteen. Hieman vastaavanlainen värähtelyspektroskopian piiriin kuuluva Raman-spektroskopia perustuu valon sirontaan näytteen molekyyleistä absorptioon sijaan.

Kuten kaikkeen mittaamiseen, myös IRS:an liittyy häiriöitä ja epävarmuustekijöitä. Yleisimpiä virhelähteitä ovat (näytteestä riippuen) erilaiset valon heijastumisesta ja sironnasta johtuvat artefaktit [3]-[13]. Erityisen ongelmalliseksi on osoittautunut nk. Mie-sironta (lähteestä riippuen joskus puhutaan myös resonantista Mie-sironnasta, RMieS), jota syntyy näytteen ollessa geometrisesti pallomainen/sylinterimäinen, ja kooltaan samaa suuruusluokkaa IR:n aallonpituuden kanssa, esimerkiksi solut ja jotkut soluelimet [4-5], [8-9], [14]. Mie-sironta näkyy erityisesti spektrin pohjaviivan epälineaarisenä vaihteluna.

Väestön ikääntyminen kehittyneissä maissa ja siihen liittyvien sairauksien hoidon ja diagnostiikan kustannusten nopea kasvu on asettanut terveydenhuoltosektorille paineita ennaltaehkäisevän toiminnan ja diagnostiikan parantamiseen. Nivelrikko on yksi yleisimpiä tuki- ja liikuntaelinsairauksia, mutta sen alkuvaiheen diagnostiikka on osoittautunut erittäin haasteelliseksi. Sairauden etenemis- ja syntymekanismit ovat edelleen osin tuntemattomia, jolloin hoito keskittyy lähinnä oireiden hallintaan. Nivelruston molekyyli rakenne on kuitenkin hyvin tunnettu, minkä vuoksi IRS:lla on pystytty tutkimaan nivelruston rakenteellisia muutoksia sairauden eri vaiheissa [1], [2]. IRS voisikin olla yksi tulevaisuuden mahdollisista nivelruston diagnostiikkavälineistä, mutta ainakin vielä käytännön toteutus on työlästä, koska IRS:a varten tarvitaan joko biopsianäyte nivelrustokudoksesta tai mittaus on tehtävä tähystyksen aikana [2]. Tässä

tutkimuksessa ei ole tarkoitus perehtyä nivelrikon mekanismeihin tai diagnostiikkaan, mutta on hyvä tiedostaa nivelrikon merkityksellisyys terveydenhuollossa ja miksi IRS voisi mahdollisesti olla yksi tulevaisuuden apuvälineistä nivelrikon diagnostiikassa.

EMSC (extended multiplicative signal correction) on IRS:ssa laajasti käytetty esikäsittelymetodi [6], [7], [14]. EMSC on suora laajennus jo 1980-luvulla kehitetystä MSC (multiplicative scatter correction) -algoritmista. EMSC-menetelmä perustuu Beer-Lambertin lakiin ja ennakkotietoihin nivelrustospektrissä esiintyvien virheiden matemaattisesta mallinnuksesta [3], [7], [12]. EMSC-mallille on kehitetty erilaisia muotoja eri tutkimusten vaatimuksia mukaillen, mutta yksi pääaiheista on ollut Mie-sironnan poistaminen. Tässä tutkielmassa pyritään Mie-sirontaa approksimoimalla rakentamaan ei-iteratiivinen ketterä EMSC-malli, jonka toimivuutta testataan sekä simuloidun nivelrustospektrin että oikean FTIR-kuvatun nivelrustospektrin esikäsittelyssä.

## 2 Tutkimuksen tausta

### 2.1 Infrapunaspektroskopia ja EMSC-malli

Infrapunaspektroskopia perustuu empiirisesti Beer-Lambertin lakiin, jonka mukaan infrapunavalon vaimeneminen riippuu tutkittavan näytteen ominaisuuksista. Spektroskopiassa on yleistä käyttää aallonpituuden sijaan aaltolukua, joka on aallonpituuden käänteisluku:

$$\tilde{\nu} = \frac{1}{\lambda} = [m^{-1}] . \quad (1)$$

Nyt tutkittavan näytteen läpäisevyydelle voidaan kirjoittaa

$$T(\tilde{\nu}) = \frac{I(\tilde{\nu})}{I_0(\tilde{\nu})} = e^{-\epsilon(\tilde{\nu})cl} . \quad (2)$$

Läpäisevyys  $T(\tilde{\nu})$  riippuu siis suoraan näytteeseen tulevan valon intensiteetin  $I_0(\tilde{\nu})$  ja näytteen läpäisseen valon intensiteetin  $I(\tilde{\nu})$  suhteesta. Voidaan vielä määritellä tämä suhde riippuvaiseksi tietylle aineelle ominaisesta absorptiotekijästä  $\epsilon(\tilde{\nu})$ , valon kulkemasta matkasta näytteen sisällä  $l$  ja mitattavan aineen konsentraatiosta  $c$ . Yleensä ollaan kuitenkin kiinnostuneita nimenomaan absorptiosta, joka voidaan ilmaista läpäisevyyden avulla

$$A(\tilde{\nu}) = -\log_{10}(T(\tilde{\nu})) = \epsilon(\tilde{\nu})cl . \quad (3)$$

Beer-Lambertin laki pätee kuitenkin täysin vain ns. ideaalitapauksessa, jolloin tutkittava näyte on homogeeninen eikä valon sirontaa tapahdu.

Beer-Lambertin lain pohjalta voidaan kuitenkin johtaa perusmuotoisen MSC-mallin yhtälö [3], [11], [12], kun oletetaan, että mitatut spektrit vastaavat muodoltaan läheisesti keskiarvospektriä:

$$A(\tilde{\nu}) = a + \bar{x}(\tilde{\nu}) \cdot b + e(\tilde{\nu}) . \quad (4)$$

MSC-mallin mukaan mitatut spektrit  $A(\tilde{\nu})$  voidaan esittää parametrien  $a$  ja  $b$ , keskiarvospektrin  $\bar{x}(\tilde{\nu})$  ja mallintamattoman residuaalin  $e(\tilde{\nu})$  avulla. Parametrit ovat

ennestään tuntemattomia, mutta niitä voidaan arvioida esimerkiksi pienimmän neliösumman menetelmällä (PNS), jolloin korjatuiksi spektreiksi saadaan

$$A(\tilde{\nu})_{korj} = \frac{A(\tilde{\nu}) - a}{b} . \quad (5)$$

Tätä MSC-mallin perusmuotoa on myöhemmin laajennettu tämänkin tutkielman perustana olevaan EMSC-malliin, jonka alunperin kehittivät Martens ja Stark [3]. EMSC-mallilla voidaan mitatuista spektreistä poistaa pohjaviivan vaihtelun epälineaarisuutta, kuten perusmuodon matemaattisesta mallista voidaan havaita:

$$A(\tilde{\nu}) = a + \bar{x}(\tilde{\nu}) \cdot b + d_1 \tilde{\nu} + d_2 \tilde{\nu}^2 + \dots + d_n \tilde{\nu}^n + e(\tilde{\nu}) . \quad (6)$$

Käytetyt termit ovat muuten samat kuin MSC-mallin perusmuodossa (yhtälö 4), mutta EMSC-malliin voidaan lisätä  $n$ :nen asteen polynomia mallintava osa. Polynomien aste voidaan valita tarpeen mukaan, ja yleensä perusmuotoon viitattaessa tarkoitetaan kvadraattisella termillä laajennettua mallia [11]. Kuten edellä MSC-mallissa, voidaan tuntemattomia parametreja estimoida PNS:n avulla, jolloin saadaan:

$$A(\tilde{\nu})_{korj} = \frac{A(\tilde{\nu}) - a - d_1 \tilde{\nu} - d_2 \tilde{\nu}^2 - \dots - d_n \tilde{\nu}^n}{b} . \quad (7)$$

Yleensä parametreilla  $a$  ja  $b$  viitataan pohjaviivan siirrokseen ja optiseen matkaan liittyvään kerrannaistekijään [3], [6], mutta periaatteessa EMSC-malli mahdollistaa muidenkin spektreihin vaikuttavien virhetekijöiden arvioimisen, jos näistä tekijöistä on tietoa *a priori* ja niitä osataan matemaattisesti mallintaa. Tätä voidaan perustella juurikin Beer-Lambertin lailla, jonka mukaan eri aineista koostuvan näytteen absorptiospektri voidaan ilmasita yksittäisten aineiden absorptioiden superpositiona [3], [11].

EMSC-malli on nykyään yleinen esikäsittelymetodi infrapunaspektroskopiassa [6], [7], [14], [15], sekä Raman-spektroskopiassa. Edellä esitetyn perusversion lisäksi on kehitetty lukuisia erilaisia versioita erilaisten tutkimusten tarpeisiin [4]-[7],[9]-[11],[16]. Erilaisten virheiden mallintamisen lisäksi sillä on pystytty sovittamaan dataa vertailukelpoiseksi eri tutkimustenkin välillä [16].

## 2.2 Mie-sironta

Valon sironta on yleinen häiriönlähde IRS:ssa [3]-[6], [9], joka vaikeuttaa kemiallisen informaation tulkintaa ja jatkoanalyysia absorptiospektreistä. Erityisesti biologisten kudosten kuvantamisessa solut ja jotkut soluelimet aiheuttavat Mie-sironnaksi kutsuttua valon siroamista [4], [6], [9]. Näytteestä sironnut valo ei välttämättä osu detektoriin, mikä ilmenee ylimääräisenä absorptiona spektreissä. Ongelmaksi muodostuu tällöin erottaa, mikä osuus absorptiosta on näytteen oikeata absorptiota ja mikä osuus johtuu valon sironnasta.

Mie-sirontaa ilmenee, kun näytteessä on kokoluokaltaan IR-valon aallonpituutta vastaavia kappaleita, kuten yksittäiset solut tai soluelimet biologisessa kudoksessa [4], [6], [9], [15]. Mie-sironta muuttaa usein sekä absorptiopiikkien amplitudeja että paikkaa, sekä aiheuttaa erityisesti tietyillä aaltoluvuilla (amidi 1 -absorptioalue) huomattavia vääristymiä spektrissä [8]. Tämän tutkielman yksi päätavoitteista onkin yrittää poistaa Mie-sirontaa nivelrustonäytteen spektristä.

Mie-sironta on nimetty löytäjänsä Gustav Mien mukaan. Mie julkaisi omat ratkaisunsa Maxwellin yhtälöihin v. 1908 ja näille ratkaisuille käytetään vieläkin nimitystä “Mie-sironta” [4], [6]. Mien kehittämä teoriaa kuvailee tarkasti sähkömagneettisen säteilyn sirontaa pallomaisista kappaleista. Mie-sironta pystytään siis hyvin eksaktisti kuvailemaan, mutta sen käytännön mallintamisessa on ongelmia. Mien kehittämä teoria on matemaattisesti vaativa ja täten Mie-sironnan täydellinen mallintaminen onkin laskennallisesti kallista ja hidasta, tosin tietokoneiden prosessointitehon lisääntyminen ja rinnakkaislaskennan mahdollisuus ovat tehneet Mie-sironnan täydellisen mallin käyttämisestä yksinkertaisempaa [4], [6]. Silti erilaisten approksimaatioiden käyttäminen on varsin yleistä niiden yksinkertaisuuden ja riittävän tarkkuuden vuoksi. Tässä tutkielmassa rakennettu EMSC-malli käyttää myös Mie-sironnan approksimaatiota, joka esitellään hieman myöhemmin.

## 2.3 Nivelruston kuvantaminen IRS:lla

Nivelrusto on koostumukseltaan hyaliinirustoa, joka ympäröi luiden nivelenpuoleisia päitä suojaten luita rasitukselta ja muodostaen lähes kitkattoman pinnan luiden välille mahdollistaen sujuvan liikkeen [1], [2], [15]. Suurin osa nivelrustosta on vettä, mutta



veden osuus vaihtelee kudoksen syvyyssuunnassa; mitä syvemmälle mennään, sen vähemmän vettä. Nivelrikossa veden osuus nivelrustossa saattaa lisääntyä, ja veden määrän muutosta onkin tutkittu potentiaalisena nivelrikon indikaattorina [2]. Nivelruston kiinteä osa koostuu enimmäkseen tyypin II kollageenista, proteoglykaanista ja muista proteiineista [1], [15]. Yleisin proteoglykaani nivelrustossa on aggregaani, joka koostuu enimmäkseen kondroitiinisulfaatista (CS1). Kuten veden osuus, myös kiinteän matriksin koostumus vaihtelee syvyyden mukaan. Lisäksi kollageenin säikeiden orientaatio muuttuu kudoksen syvyyden mukaan mahdollistaen osaltaan eri kerroksien erilaiset biomekaaniset ominaisuudet [1].

IRS soveltuu menetelmänä hyvin nivelruston tutkimiseen, koska nivelruston kemiallinen koostumus on jo hyvin tunnettu, eikä IRS vahingoita tai aiheuta rakenteellisia muutoksia näytteeseen kuvantamisen aikana. Koska terveen nivelruston rakenne tunnetaan, tiedetään myös melko tarkasti millainen absorptioprofiili nivelrustolla on. Nivelruston absorptiospektriä voidaankin approksimoida hyödyntämällä aikaisempaa tietoa koostumuksesta sekä Beer-Lambertin lakia; superpositionperiaatteen nojalla voidaan nivelruston absorptiospektriä approksimoida summaamalla kollageenin ja CS:n puhdasainespektrit.

Kuten jo aikaisemmin mainittiin, biologisia näytteitä kuvattaessa Mie-sironnan esiintyminen on todennäköistä. Lisäksi biologiset näytteet ovat hyvin heterogeenisiä ja usein niiden rakenne vaihtelee syvyyden mukaan, jolloin ennakkotieto rakenteesta ja koostumuksesta on hyvin olennaista. Ylipäänsä voidaan ajatella, että esikäsittelyä tehtäessä tulos paranee mitä enemmän ja tarkemmin näytteestä ennalta tiedetään. Tämäkin tutkielma nojaa EMSC-mallin rakentamisessa vahvasti ennakkotietoihin nivelruston koostumuksesta.

### **3 Tutkimuksen tavoitteet**

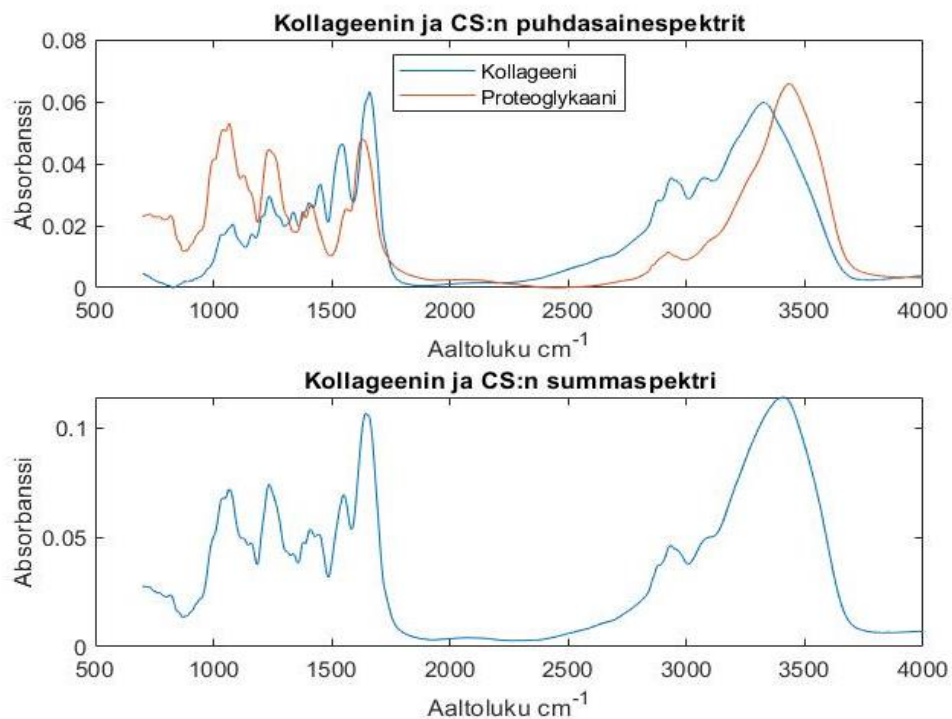
Tutkimuksen tavoitteena oli selvittää kuinka hyvin ketterällä EMSC-algoritmillä pystytään poistamaan erityisesti Mie-sirontaa FTIR-absorptiospektristä. EMSC-mallia varten Mie-sirontaa mallinnettiin matemaattisella approksimaatiolla, ja testattiin ensin simuloidulla datalla. Lopuksi verrattiin muodostetun EMSC-mallin ja täydellisen Mie-sironnan mallin sisältävän RMieS-algoritmin korjauksia oikealla mittausdatalla.

## **4 Materiaalit ja menetelmät**

Tutkimus on toteutettu kahdessa osassa; ensin on tutkittu, kuinka hyvin kehitetty EMSC-algoritmi suoriutuu osaspektreillä simuloidun nivelruston absorptiospektrin, johon on lisätty keinotekoisia virhelähteitä, esikäsittelystä. Sen jälkeen algoritmia on testattu oikealla sorkkaeläimen nivelrustonäytteestä kuvatulta FTIR-datalla, ja verrattu korjattuja spekreja monimutkaisemman korjausalgoritmin tuloksiin. Tutkimuksessa datan käsittelyyn ja algoritmien toteuttamiseen on käytetty Matlabin versiota R2018b(MathWorks Inc).

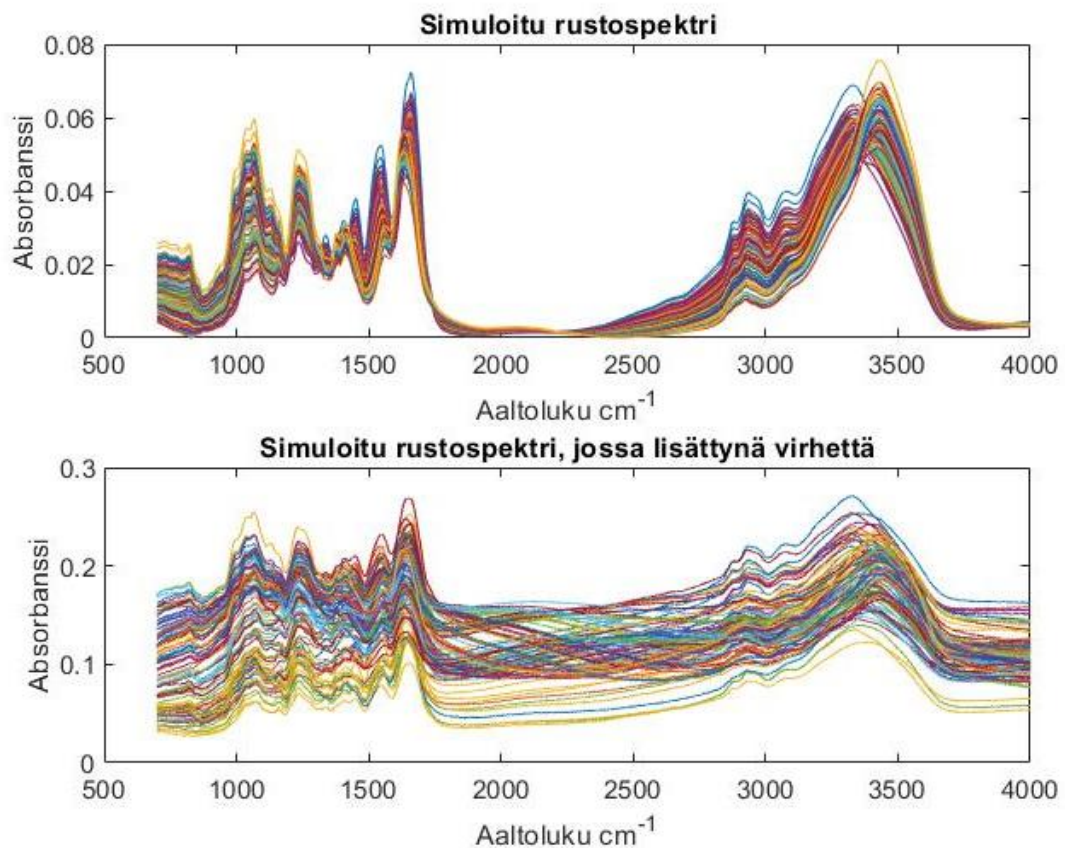
### **4.1 Osaspektreillä simuloitu absorptiospektri**

Aikaisemmissa tutkimuksissa on osoitettu, että osaspektreillä voidaan mallintaa IR-absorptiospektriä tyydyttävästi [4], [9], [14], [15]. Tutkimuksessa käytetty simuloitu nivelruston absorptiospektri on muodostettu kollageenin ja kondroitiinisulfaatin (CS) puhdasainespektrien summaspektrinä. CS muodostaa suurimman osan nivelruston proteoglykaanista, ja kollageeni on nivelruston kiinteän matriksin yleisin molekyyli. Spekitrit ovat esitetty Kuvassa 1.



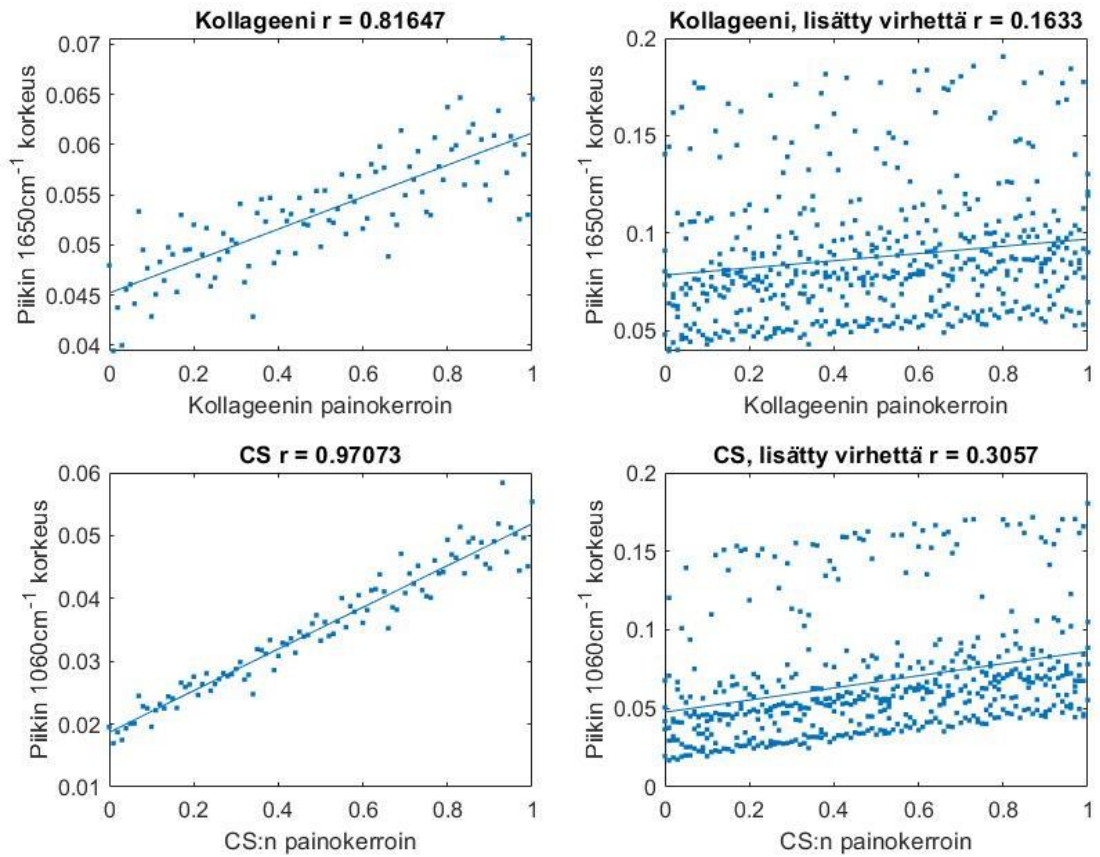
**Kuva 1.** Kollageenin ja CS:n puhtasainespektrit, sekä niiden summaspektrinä muodostettu nivelrustospektri.

Antamalla CS:lle ja kollageenille satunnaistetusti eri painokertoimia on simuloitu nivelrustokudoksen heterogeenisyyttä eri syvyyksillä ja muodostettu useita rustospektrejä sisältävä datasetti. Datasettiin on vielä lisätty pohjaviivan vaihtelua (ensimmäisen ja toisen asteen), satunnaiskohinaa, kerrannaisvirhettä sekä approksimoitua Mie-sirontaa, jotta spektrit saataisiin muistuttamaan oikeata mittausdataa. Kuvassa 2 on esitetty simuloidut spektrit ilman virheitä ja virheiden lisäämisen jälkeen.



**Kuva 2.** Simuloitu rustospektri ennen ja jälkeen virhetermien lisäystä.

Ennen esikäsittelyä on vielä todettu, että CS:n ja kollageenin painokertoimet korreloivat lineaarisesti muodostettujen (vain kohinaa sisältävien) spektrien vastaavien absorptiopiikkien kanssa ( $r_{\text{kollageeni}} = 0,82$  ja  $r_{\text{CS}} = 0,97$ ). Kuvassa 3 on havainnollistettu, että korrelaatio häviää verrattaessa virheitä sisältävien spektrien vastaavia absorptiopiikkejä kollageenin ja CS:n painokertoimiin ( $r_{\text{kollageeni}} = 0,16$  ja  $r_{\text{CS}} = 0,31$ ).



**Kuva 3.** Kollageenin ja CS:n korrelaatiot valittujen absorptiopiikkien ja painokertoimien välillä. Spektrien esikäsittelyssä on pyritty laajentamaan yhtälössä (6) esitettyä EMSC:n perusmuotoa erityisesti Mie-sironnan poistamiseksi. Kohler et al [6] on esittänyt EMSC:lle muotoa

$$A(\tilde{\nu}) = a + b \cdot m(\tilde{\nu}) + f \cdot Q(\tilde{\nu}) + e(\tilde{\nu}) . \quad (8)$$

Yhtälö kuvaa sitä, kuinka mikä tahansa spektri  $A(\tilde{\nu})$  voidaan teoriassa esittää selittävien komponenttien ja referenssispektrin summana. Tässä tapauksessa  $a$  kuvastaa pohjaviivan vaihtelua,  $b$  liittyy kerrannaistekijään, joka muuttaa mm. piikkien korkeutta ja  $m(\tilde{\nu})$  on valittu referenssispektri aaltoluvun funktiona. Mie-sironnan mallinnukseen pureudutaan tarkemmin myöhempanä, tässä vaiheessa tyydytään kuvamaan Mie-sironnan osuutta komponenteilla  $f$  ja  $Q(\tilde{\nu})$ . Lopuksi eri aaltoluvuilla esiintyvät muut mallintamattomat tekijät esitetään residuaalina  $e(\tilde{\nu})$ .

Mikäli jokaisen spektrin Mie-sirontaa  $Q(\tilde{\nu})$  pystytään mallintamaan, voidaan pienimmän neliösumman (PNS) menetelmällä arvioida edellä esitettyjä kertoimia, ja korjatuille spektreille muodostaa yhtälö

$$A_{korjattu}(\tilde{\nu}) = \frac{A(\tilde{\nu}) - a - f \cdot Q(\tilde{\nu})}{b} . \quad (9)$$

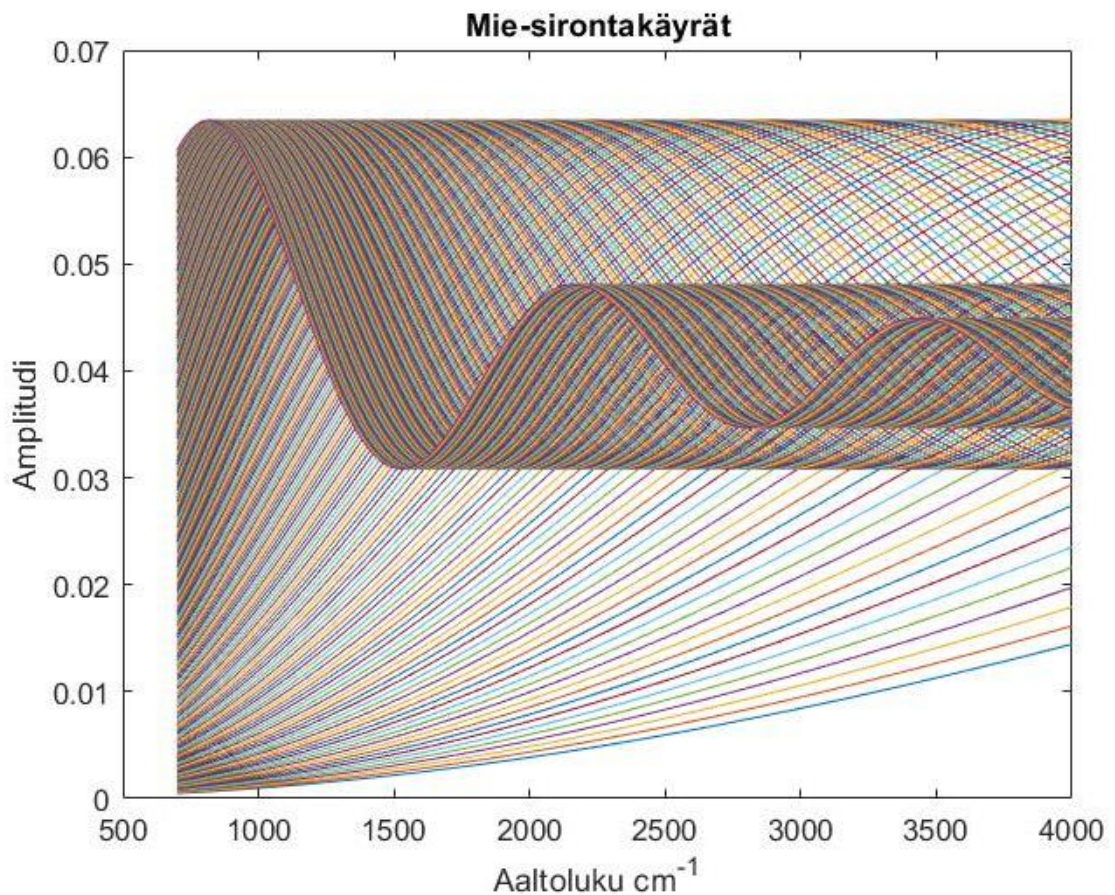
Mie-sironnan tarkka mallintaminen on kuitenkin matemaattisesti haastavaa ja laskennallisesti kallista, joten tietyissä tapauksissa on järkevää käyttää tarkan mallintamiseen sijaan approksimaatiota. Yksi tällaisista approksimaatioista on van de Hulstin esittämä approksimaatio, joka tietyin rajoittein mallintaa Mie-sirontaa varsin tarkasti. Approksimaatio voidaan esittää muodossa

$$Q(\lambda) = 2 - \left(\frac{4}{\rho}\right) \sin \rho + \left(\frac{4}{\rho^2}\right) (1 - \cos \rho), \text{ missä} \quad (10)$$

$$\rho(\lambda) = \frac{4\pi a(n-1)}{\lambda} . \quad (11)$$

On syytä huomioda, että tämä yhtälö on kirjoitettu aallonpituuden, eikä aaltoluvun, funktiona. Yhtälössä  $a$  on pallon halkaisija ja  $n$  on pallon sisä- ja ulkpuolen taittokertoimien suhde. Tällä approksimaatiolla voidaan tietyin oletuksin päästä n. 1% päähän Mie-sironnan täydellisen mallin tuottamista tuloksista [6], [15]. Kuvassa 4 on esitettyä yhtälön (10) perusteella vaihtelevilla halkaisijan arvoilla muodostettua Mie-sirontaa, jota on lisätty simuloituun rustospektriin.





**Kuva 4.** Yhtälön 9 perusteella muodostettuja approksimaatioita Mie-sirontakäyristä.

Jotta PNS-menetelmällä voidaan arvioida sopivat kertoimet yhtälöön (9), meidän täytyy pystyä jakamaan kiinnostavat (häiriö)tekijät komponentteihin. Pohjavaihtelu on kohtuullisen yksinkertaista esittää  $n$ -asteisena polynomina, samoin kerrannaistekijän voidaan olettaa olevan muotoa  $\text{vakio} * \text{referenssispektri}$ . Referenssispektri on tässä tapauksessa valittu yksi simuloituista rustospektreistä. Yleensä ennalta ei ole tiedossa, miltä “täydellinen” spektri näyttäisi, vaan referenssinä käytetään esimerkiksi keskiarvospektriä tai silmämääräisesti vähän häiriötä sisältävää spektriä [3], [6], [8], [10], [13]. Mie-sironnan komponenttien arviointi ei kuitenkaan ole aivan yhtä suoraviivaista. Pääkomponenttianalyysillä (PCA) voidaan kuitenkin nimensä mukaisesti arvioida Mie-sironnan pääkomponentteja jokaisessa spektrissä [4]-[6]. Simuloitujen spektrien tapauksissa kuusi ensimmäistä pääkomponenttia selittää  $>95,5\%$  Mie-sironnan varianssista. Koska spektrien Mie-sirontaa voidaan mallintaa seuraavasti



$$\sum_{k=1}^A g_k \cdot p_k(\tilde{\nu}), \text{ missä} \quad (12)$$

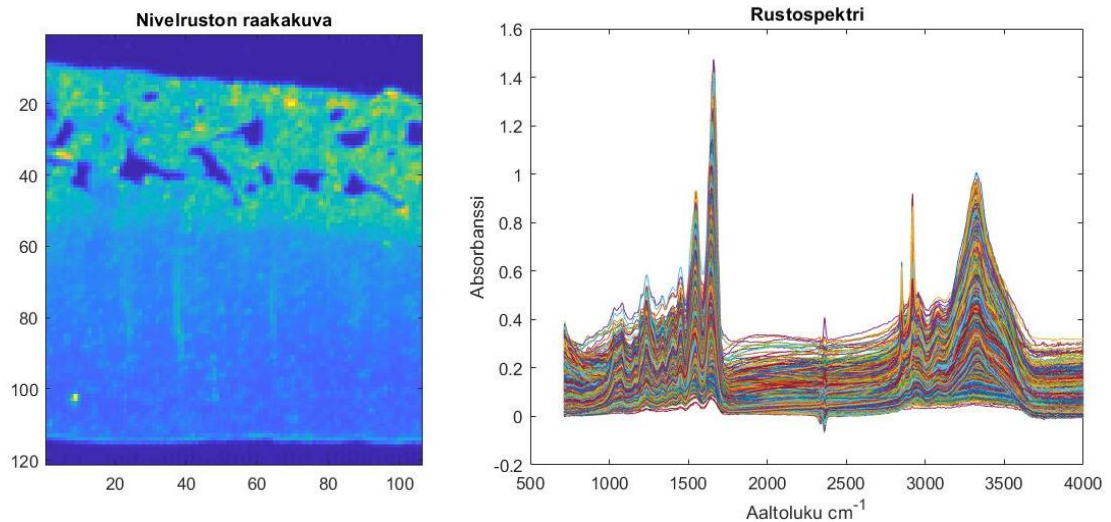
$p_k$  on PCA:lla saatu sironnan komponenttien painotus ja  $g_k$  komponenttiin liittyvää kerroin, voidaan nyt PNS-menetelmällä ratkaista kertoimet  $g_k$  painotuksien avulla. Nyt pystytään vihdoinkin ratkaisemaan yhtälön (8) muuttujat  $a$ ,  $b$ , ja  $f$  ja muodostamaan varsinainen EMSC-algoritmi spektrien esikäsittelyä varten. Mallia varten on estimoitu vakiokertoimista pohjaviivan vaihtelua, kerrannaistekijää ja Mie-sirontaa. EMSC-malliin voidaan vielä lisäksi lisätä painokerroinvektori, kun PNS:llä arvioidaan haluttuja komponentteja. Painokertoimien tarkoituksena on pienentää EMSC-mallin korjausta kiinnostavilla alueilla, jotta mahdollisimman suuri osa kemiallisesta informaatiosta säilyisi. Esimerkiksi jos simuloidun rustospektrin tapauksessa olisimme kiinnostuneita näytteen kollageenimäärästä ja -koostumuksesta, voisimme antaa aaltoluvulle 1650 ja sen ympäristölle pienemmän painokertoimen. Painokertoimien oikeaoppinen valinta vaatii käyttäjältä ennakkotietoa näytteen kiinnostavien alueiden aaltoluvuista. Simuloitujen spektrien tapauksessa painokerroinvektori on valittu siten, että aaltolukualueille 900-1800 ja 2600-3700 on annettu painokerroin 0.3.

Jotta korjauksen onnistuneisuus voidaan todentaa, voidaan tarkastella esimerkiksi korjauksen hyvyyttä epäkiinnostavilla alueilla, ts. valitaan spektristä alue, jonka ei tulisi sisältää kemiallista informaatiota, ja tulisi näin ollen olla hyvin lähellä nollaa. Tämä voidaan laskea neliöllisellä keskiarvolla (RMS); mitä lähemmäksi nollaa RMS-arvot saadaan (ideaalitapauksessa  $\text{RMS} = 0$  kauttaaltaan), sitä vähemmän variaatiota kyseisellä alueella on korjauksen jälkeen. Vertaamalla RMS-arvoja korjaamattomien ja korjattujen spektrien välillä, saadaan jonkinlainen kuva algoritmin hyvyydestä. Ennen korjausta oli määritetty korrelaatio kollageenin ja kondroitiinin painokertoimien, ja vastaavien absorptiopiikkien välillä. Todettiin, että virhetermien lisäys poisti korrelaation, joten tarkastellaan säilyykö korrelaatio esikäsittelyn jälkeen.

## 4.2 FTIR-nivelrustospektri

Vaikka edellä käsiteltiin simuloituun nivelrustospektriin olikin lisätty yleisimpiä virhetermejä, on syytä tutkia EMSC-mallin toimivuutta oikealla mittausdatalla, jotta saadaan luotettava arvio algoritmin korjauskyvystä. Käytetty mittausdata on peräisin

FTIR-mikroskoopilla kuvannetuista sorkkaeläimen nivelrustonäytteistä. Kuvassa 5 on esitetty raakakuva näytteestä, josta voidaan havaita ns. tyhjää tilaa yläreunassa. Ennen varsinaista esikäsittelyä poistetaan datasta tyhjät alueet ja muotoillaan data 2D-muotoon. Mitatut absorptiospektrit on esitetty niin ikään Kuvassa 5.



**Kuva 5.** Nivelruston raakakuva ja absorptiospektri.

Rakennetun EMSC-mallin korjauskykyä oikealla mittausdatalla verrataan P. Bassanin kehittämään täydellisempään EMSC-algoritmiin RMieS\_EMSC\_v5 (viitataan RMies-mallina/algoritmina jatkossa) [17]. Tämä RMieS-malli sisältää Mie-sironnan arvioinnin täydellisen Mie-teorian avulla, ja on lisäksi iteratiivinen, toisin kuin vain kerran datan läpikäyvä EMSC-mallimme. Esikäsittely on RMieS-mallissa toteutettu kolmella iteraatiolla, referenssispektrinä on käytetty Matrigelin absorptiospektriä ja pääkomponentteja on laskettu kahdeksan. Vertaantuvuuden vuoksi myös EMSC-malliin on tässä vaiheessa lisätty RMieS-mallissa esiintyvä pohjaviivan aaltoluvun funktiona tapahtuva lineaarisen vaihtelun estimointi. Referenssispektri on EMSC-malliin valittu simuloitu rustospektri, jonka koostumukseksi on asetettu 70% kollageenia ja 30% CS:a. Lisäksi painokerroinvektori on valittu siten, että amidi- ja hiilidioksidipiikkien, jotka esiintyvät n. 1500-2000 ja 2300-2400 aaltoluvuilla, alueilla korjauskerroin on vain 0.01, kuten RMieS-mallissa. Aikaisemmissa tutkimuksissa on todettu juuri näiden aaltolukualueiden sisältävän informaatiota nivelruston kollageeni- ja CS-määristä [1],

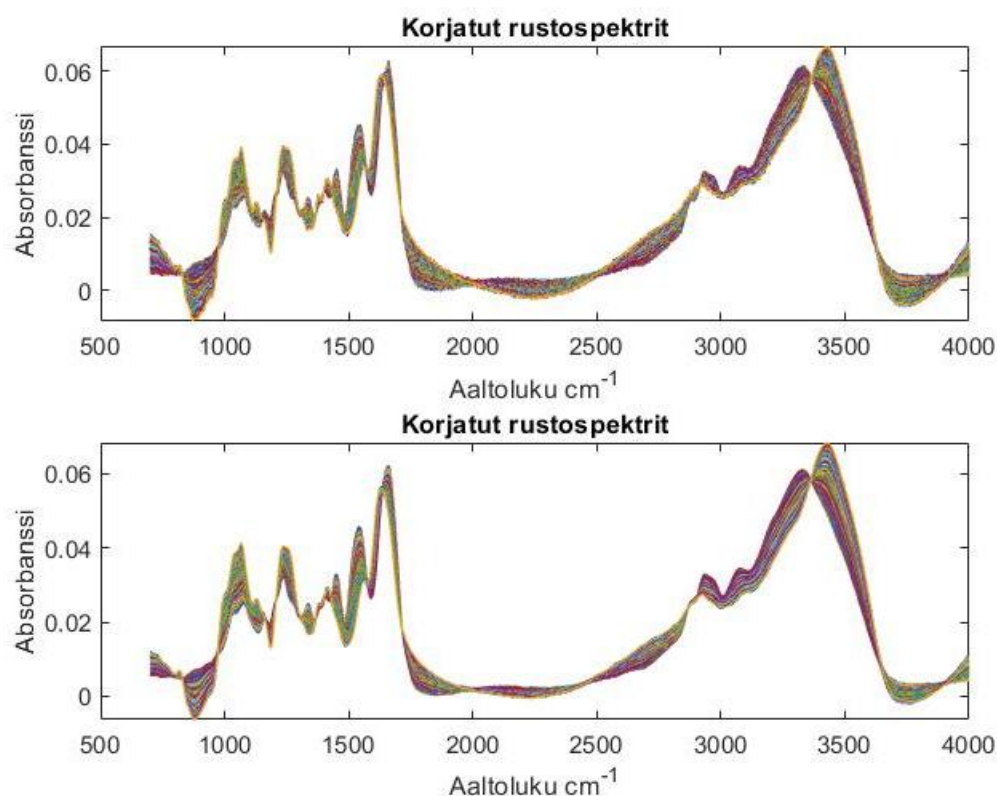
[15]. EMSC-mallin korjausta ei voida oikean mittausdatan tapauksessa verrata ihannespektriin, koska sellaista yksikäsitteistä virheetöntä nivelrustospektriä, joka täydellisesti mallintaisi ruston kemiallista koostumusta, ei ole olemassakaan. Tässä tutkimuksessa tehdään oletus, että kehitetty RMieS-malli tuottaa riittävän hyvän korjauksen, johon EMSC-mallin korjausta voidaan verrata.

Nyt ei myöskään ole saatavilla painokertoimia kollageenin ja CS:n pitoisuuksille, eikä absorptiopiikkien tarkkaa paikkaa ole helppoa määrittää mittausdatasta. EMSC-mallin korjauksen hyvyttä voidaan tässä tapauksessa arvioida suoraan laskemalla korrelaatio RMieS-mallin tuottaman korjauksen kanssa. Näiden kahden mallin välillä voidaan myös laskea neliöllisen keskiarvon virhe (RMSE), joka kertoo kuinka lähellä toisiaan mallin spektrit ovat eri aaltoluvuilla. Lopuksi verrataan vielä, minkälainen vaikutus painokerroinvektorilla ja suodatuksella on spektrien esikäsittelyyn.

## **5 Tulokset**

### **5.1 Osaspektreillä simuloitu absorptiospektri**

EMSC-mallilla korjatut simuloitut rustospektrit on esitetty Kuvassa 6. Kuten huomataan, on korjaus ainakin päällisin puolin onnistunut. Lisäksi Kuvassa 6 on esitetty myös EMSC-mallin korjaus painokertoimilla ja suodatuksella näiden vaikutuksen havainnollistamiseksi. Pelkkä silmämääräinen tarkastelu ei kuitenkaan anna riittävän hyvää arviota korjauksen onnistuneisuudesta, joten Kuvassa 7 on esitetty kollageenin ja CS:n painokertoimien ja vastaavien absorptiopiikkien välinen korrelaatio korjauksen jälkeen. Aikaisemmin todettiin, että virheitä sisältävissä spektreissä tämä korrelaatio hävisi, mutta EMSC-mallin korjauksen jälkeen vahva korrelaatio on jälleen havaittavissa. Myös Taulukossa 1 esitetyt RMS-arvot ovat korjauksen jälkeen selvästi pienempiä, mikä viittaa siihen, että esikäsittely EMSC-mallilla oli onnistunut. On syytä mainita, että simuloitujen rustospektrien tapauksessa sekä painotuksella että suodatuksella ei ollut merkittävää vaikutusta korrelaatio- tai RMS-arvoihin.



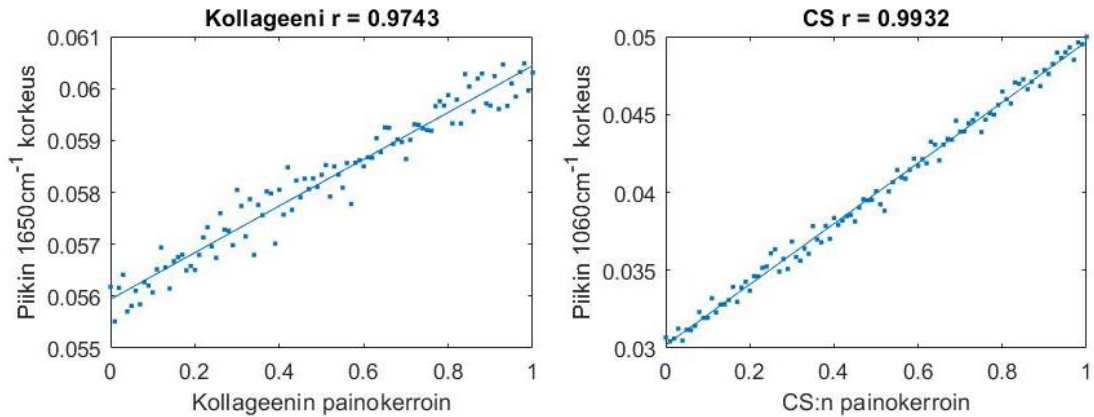
**Kuva 6.** EMSC-mallilla korjatut rustospektrit. Alemmassa kuvaajassa on EMSC-mallissa käytetty painokerroinvektoria jonka jälkeen on tehty Savitzky-Golay -suodatus.

**Taulukko 1.** Valitulla aaltolukualueella lasketut RMS-arvot simuloituille rustospektreille.

$\tilde{\nu} = \text{n. } 1800\text{-}2200$	RMS	RMS max
Korjaamaton	0.1117	0.1612
Korjattu	0.0022	0.0031

Vaikka simuloitun rustospektrin esikäsittely antoi rohkaisevia tuloksia EMSC-mallin toimintaan liittyen, ei tästä voida vielä vetää vahvoja johtopäätöksiä. Siitä huolimatta, että simuloituun rustospektriin olikin lisätty yleisesti mittauksissa esiintyviä virheitä, ei tilanne silti vastaa oikean mittausdatan monimutkaisuutta. Tilanteen voidaan ajatella olleen luonteeltaan deterministinen; koska tiedettiin täsmälleen, mitä virheitä ja miten mallinnettuna spektriin oltiin lisätty, voitiin esikäsittely suunnitella juuri näitä virheitä

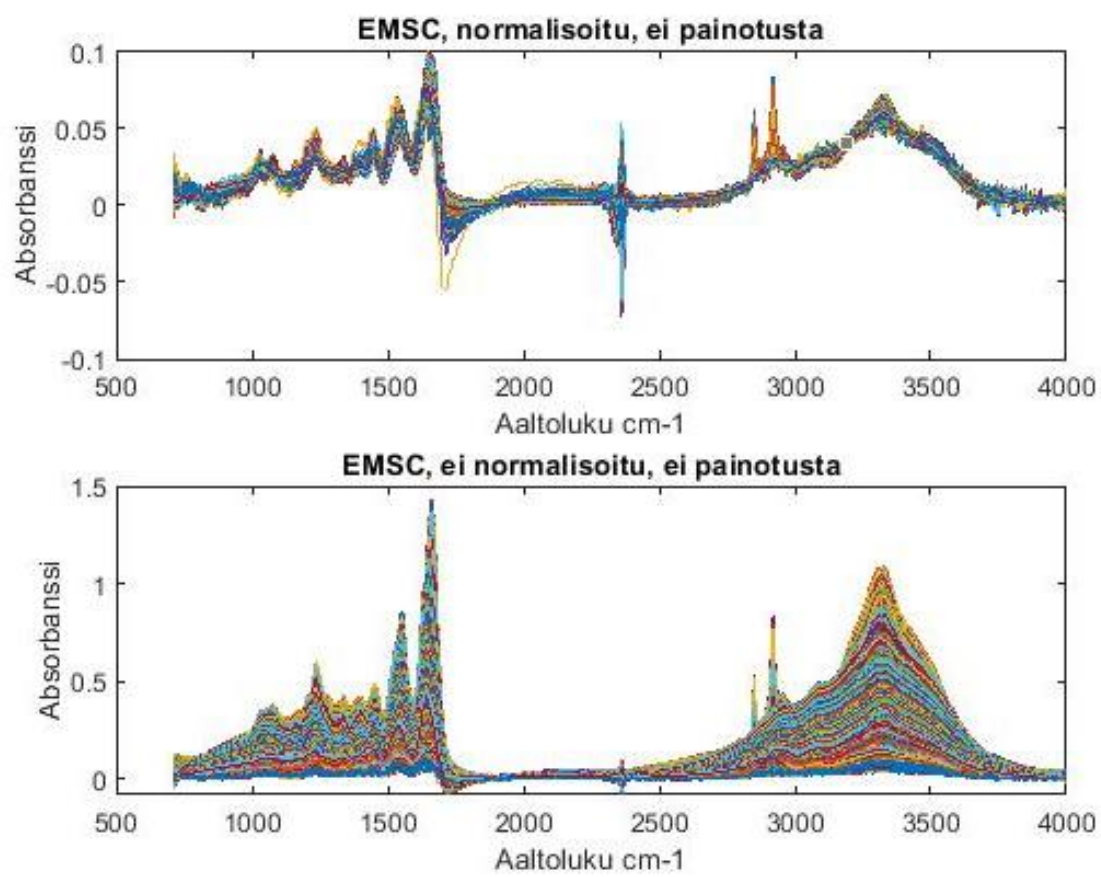
vastaamaan. Tällä perusteella kehitettyä EMSC-mallia ei siis vielä voi pitää yleispätevästi toimivana.



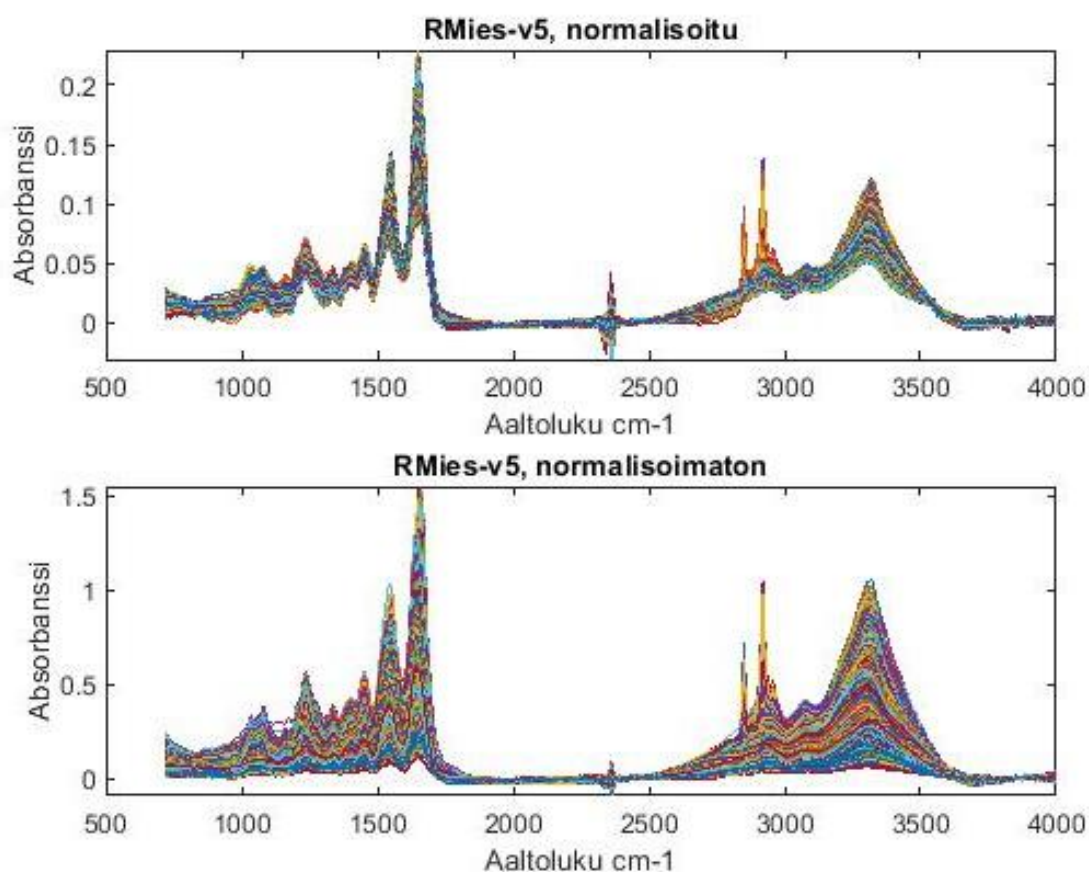
**Kuva 7.** EMSC-mallilla korjattujen spektrien absorptiopiikkien ja painokertoimien välinen korrelaatio.

## 5.1 FTIR nivelrustospektri

Nivelrustonäytteestä saadut absorptiospektrit esikäsiteltiin samalla EMSC-mallilla kuin simuloidut rustospektrit. Malliin oli ainoastaan lisätty luvussa 4.2 mainittu lineaarinen pohjaviivan vaihtelu sekä valittu erilainen painokerroinvektori. Kuvassa 8 on esitetty normalisoidut ja normalisoimattomat nivelrustospektrit EMSC-mallin korjauksen jälkeen. Tässä vaiheessa esikäsittely on tehty vielä ilman painokerroinvektoria. Välittömästi huomataan, että osassa normalisoiduista spektreistä on selkeätä ”overshootia”, eli arvot käyvät hetkellisesti todella suurina tai pieninä. Ilmiö voi selittyä sillä, että spektrit joissa ylikompensaatiota esiintyy ovat rustonäytteen reunoilta. Näytteen reunalla signaali saattaa olla heikko, sirontaa saattaa olla runsaasti tai näytteen koostumus selvästi erilainen kuin keskempänä, jolloin absorptiospektrin muotokin muuttuu. Koska nämä spektrit poikkeavat muodoltaan suuresti valitusta referenssispektristä, tapahtuu EMSC-mallin korjauksessa ylikompensaatiota. Toisaalta Kuvassa 9 esitetyissä RMies-mallin korjauksissa tai normalisoimattomissa spektreissä vastaavaa ylikompensaatiota ei esiinny.



**Kuva 8.** EMSC-mallilla korjatut FTIR-nivelrustospektrit.

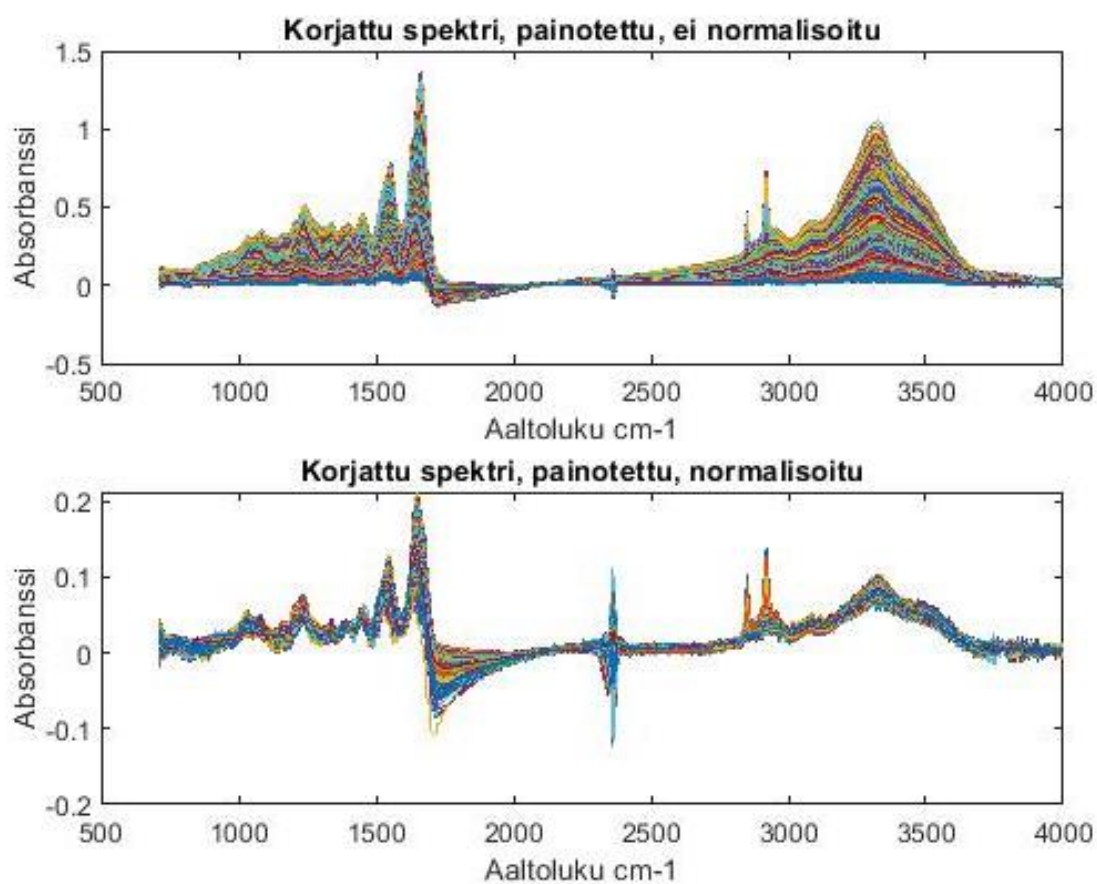


**Kuva 9.** RMies-mallilla korjatut FTIR-nivelrustospektrit.

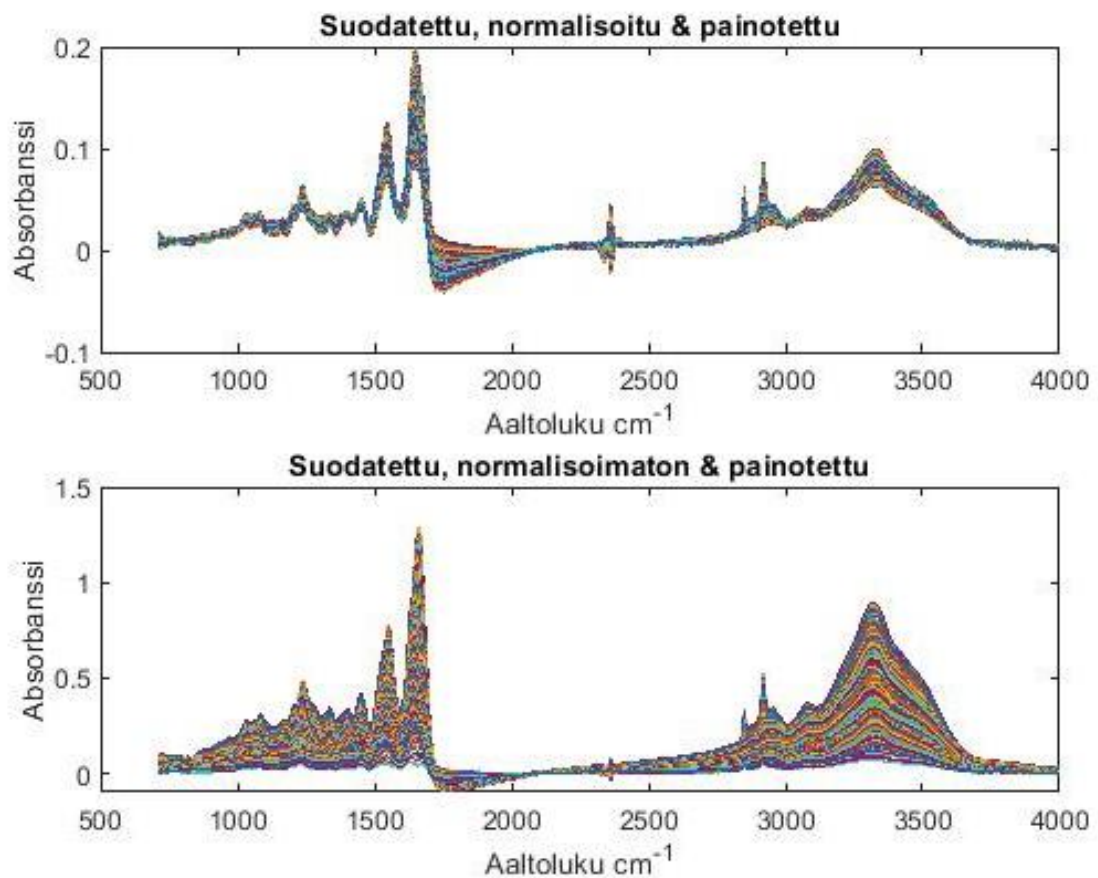
EMSC-mallin korjausta voidaan koittaa vielä parantaa painokerroinvektorin ja suodatuksen avulla. Suodattimena on käytetty toisen asteen Savitzky-Golay -suodatinta ikkunan leveydellä 17. Parametrit ovat empiirisesti valittuja. Kuvassa 10 ja 11 on esitetty, miten pelkkä painokerroin, sekä painokerroin ja suodatus vaikuttavat normalisoitujen ja normalisoimattomien spektrien korjaukseen. Painokerroinvektorin käyttö ei näytä poistavan ylikompensaatiota, mutta toisaalta spektrien informaatio ainakin amplitudien osalta säilyy normalisoiduissa spektreissä paremmin. Suodatus näyttää kuitenkin poistavan ylikompensaation spektreistä, sekä silmämääräisesti huomattavasti vähentävän variaatiota eri spektrien välillä. Aluksi tämä voi vaikuttaa hyvältä tulokselta, mutta on syytä muistaa, mikä on esikäsittelyn tavoite, ja mitä informaatiota alkuperäinen mittausdata sisältää. Nyt vaikuttaa siltä, että suodatus on yhdentänyt spektrejä liikaa. Koska spektrit ovat eri osista rustonäytettä, on suodatus todennäköisesti johtanut



kemiallisen informaation häviämiseen variaatiota vähentämällä. Jatkoanalyysi tällaisesta datasta voisi olla jokseenkin mahdotonta, eikä esikäsittely olisi palvellut tarkoitustaan.



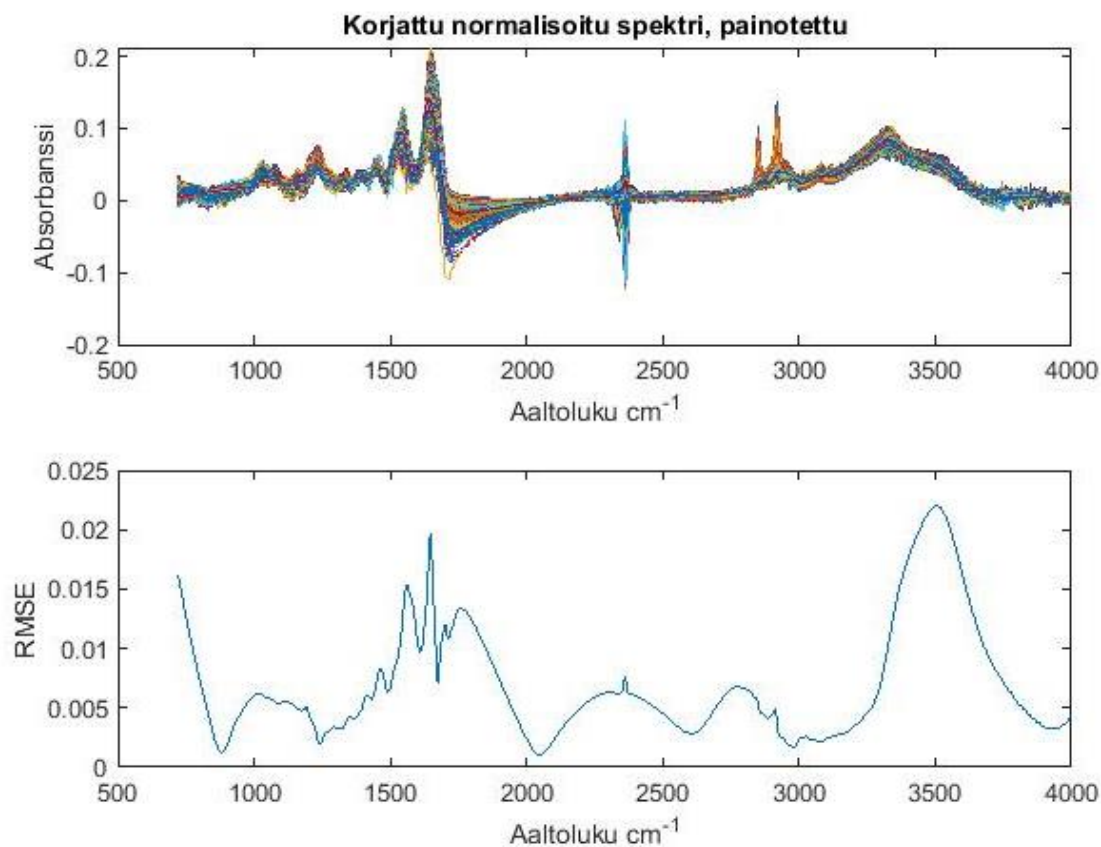
**Kuva 10.** EMSC-mallin mallin korjaus painokerroinvektorilla.



**Kuva 11.** EMSC-mallin korjaus painokerroinvektorilla ja suodatuksella.

Kuten aiemmin mainittu, ei silmämääräinen arvio vielä välttämättä kerro mitään esikäsittelyn hyvyydestä. Laskemalla RMSE-arvot korjausten välille voidaan luotettavammin arvioida korjausten onnistuneisuutta. Kuvassa 12 on lisäksi havainnollistettu, miten RMSE-arvot vaihtelevat aaltoluvun mukaan. Odotetusti RMSE-arvojen maksimit saavutetaan korkeimpien absorptiopiikkien kohdalla, missä variaatio on suurimmillaan. Taulukossa 2 on laskettuna RMSE-arvojen keski- ja maksimi-arvot eri korjausten kesken. Arvot näyttävät vahvistavan oletuksia suodattamisen hyödystä; RMSE-arvot eivät ole pienempiä suodatetuissa signaaleissa, vaikka ne päällisin puolin vaikuttivat hyviltä. Tämä tukee oletusta siitä, että suodatuksen takia on menetetty alkuperäisen datan sisältämää informaatiota. Toisin kuin simuloitujen rustospektrien tapauksessa, näyttää painotuksella olevan merkittävä positiivinen vaikutus korjausten

hyvyyteen RMSE-arvojen valossa niin normalisoiduissa kuin normalisoimattomissa spektreissä.



**Kuva 12.** Havainnollistus, kuinka RMSE-arvot vaihtelevat eri aaltoluvuilla.

**Taulukko 2.** RMSE-arvot eri esikäsittelykonfiguraatioilla.

EMSC-korjaus	RMSE (keskiarvo, maksimi)	
	Suodattamaton	Suodatettu
Normalisoitu, ei painotusta	0.0095, 0.0870	0.0095, 0.0873
Normalisoitu, painotettu	0.0069, 0.0220	0.0070, 0.0220
Normalisoimaton, ei painotusta	0.0752, 0.2332	0.0759, 0.2286
Normalisoimaton, painotettu	0.0462, 0.1977	0.0486, 0.2281

Taulukkoon 3 on laskettu vielä korrelaatiot EMSC- ja RMies-mallin korjauksien välille. Arvot ovat hyvin linjassa laskettujen RMSE-arvojen kanssa; painotus näyttää parantavan korrelaatiota sekä normalisoiduissa että normalisoimattomissa spektreissä. Suodatuksella ei tässäkään tapauksessa näytä olevan positiivista vaikutusta esikäsittelyyn.

**Taulukko 3.** Korrelaatio EMSC- ja RMies-mallin korjauksien välillä.

EMSC-korjaus	Korrelaatio	
	Suodattamaton	Suodatettu
Normalisoitu, ei painotusta	0.8971	0.8964
Normalisoitu, painotettu	0.9516	0.9514
Normalisoimaton, ei painotusta	0.8734	0.8639
Normalisoimaton, painotettu	0.9388	0.9308

Vaikka tutkimuksen päätavoitteisiin ei kuulunut EMSC-mallin suorituskyvyn optimointi tai arviointi, eikä Matlab muutenkaan tehokkuutensa puolesta sovellu todella isojen datasettien käsittelyyn [5], on silti hyödyllistä verrata kehitettyä mallia täydellisempään RMies-malliin, jotta voidaan ymmärtää syitä, miksi käytetyn EMSC-mallin tuottama korjaus ei ole aivan yhtä hyvä. EMSC-mallilla nivelrustonäytteen spektrisetin esikäsittely kesti n. 36 sekuntia, kun vastaava aika RMies-mallilla oli kolmella iteraatiolla n. 2,5 tuntia. Taulukosta 4 käy ilmi, että aikakompleksisuudeltaan rakennettu EMSC-malli on luokkaa  $O(n)$ , kun datan muokkaamista sopivaan muotoon ei oteta huomioon. Suoritus aika on siis lineaarisesti riippuvainen datasetin koosta, jos lisätään ainoastaan mittausten määrää, muttei yksittäisten mittausten mittauspisteitä. Ero laskenta-ajoissa johtuu pitkälti Mie-sironnan erilaisesta mallinnuksesta ja iteratiivisuudesta.

**Taulukko 4.** EMSC-mallin suoritus aika noin kaksinkertaistuu datasetin koon kaksinkertaistuessa.

Datasetin koko	Suoritus aika (s)
5191 x 853	16,24
10382 x 853	32,02
20764 x 853	65,95

## 6 Pohdinta

Tässä tutkimuksessa pystyttiin osoittamaan, että kohtuullisen yksinkertaisella EMSC-mallilla pystytään approksimoimaan täydellisemmän mallin tuottamia tuloksia. Saatu tulos tukee myös väittämää, että EMSC-menetelmä on varsin joustava, ja sillä kyetään moninaisten virheiden poistoon esikäsittelyvaiheessa, jos virheiden matemaattinen mallinnus on mahdollista. Tulos on hyvin linjassa muun kirjallisuuden kanssa [3]-[11], [13].

Jatkotutkimuksena voisi olla hyödyllistä selvittää, kuinka hyvin käyttämämme ei-iteratiivinen EMSC-malli suoritusi esikäsittelystä, jos Mie-sironnan approksimaation sijasta käytettäisiin täydellistä matemaattista mallinnusta Mie-sironnalla. Vertailukohtana käytetty RMieS-malli sekä monet muut kehitetyt algoritmit perustuvat iteratiiviseen menetelmään, joka on laskennallisesti hitaampaa ja kalliimpaa [4], [10], toisin kuin vain kerran spektrit läpikäyvä EMSC-mallimme. Lisäksi voitaisiin vielä lähemmin tarkastella referenssispektrin ja painokerroinvektorin valinnan vaikutusta tuloksiin, koska ainakin referenssispektrin valinta on aikaisempien tutkimusten mukaan yksi kriittisimpiä vaiheita EMSC-mallin rakentamisessa [5], [9], ja on myös ehdotettu useamman referenssispektrin käyttöä [7]. Näiden ehdotusten lisäksi EMSC-mallin yleistettävyyttä tulisi tutkia tarkemmin; tässä tutkimuksessa mallia testattiin vain simuloidulla nivelrustodatalla sekä yhdellä oikealla nivelrustomittauksella, mikä on varsin suppea otos. Sisällyttämällä useamman eri nivelrustonäytteen saataisiin parempi kuva rakennetun mallin suorituskyvystä.

Aikaisemmin on myös esitetty, että mikäli esikäsittelyssä käytetään sekä EMSC-menetelmää että derivaattapohjaista korjausta, kuten käyttämämme SG-suodatus, tulisi noudattaa tiettyä järjestystä [14]. EMSC-menetelmää on tässä tutkimuksessa kuitenkin käytetty ennen SG-suodatusta, mikä saattaa osin selittää, miksi suodatksesta ei näyttänyt olevan hyötyä laskettujen arvojen perusteella. Toisaalta SG-suodatus on yleinen menetelmä spektroskopiassa, joten tässä tapauksessa saattaa olla myös kyse huonosta suodattimen parametrien valinnasta. Myös PCA:lla ratkaistujen Mie-sironnan pääkomponenttien määrää voitaisiin muuttaa; tässä tutkimuksessa sekä simuloduille että FTIR-spektreille käytettiin EMSC-mallissa kuutta pääkomponenttia, jotka selittivät n. 95% varianssista, mutta on ehdotettu, että pääkomponentteja tulisi valita niin monta, että

ne selittäisivät >99,95% varianssista [10]. Toisaalta ei voida valita myöskään mielivaltaisen montaa pääkomponenttia kohinan lisääntymisen riskin vuoksi [10], [13].

## **7 Yhteenveto**

Muutamista huomionarvoisista puutteista huolimatta, tai ehkä jopa niiden takia, pystyttiin osoittamaan, että ketterällä ja yksinkertaisella EMSC-mallilla pystytään mallintamaan ja poistamaan absorptiospektreistä matemaattisesti haastaviakin virhetekijöitä. EMSC-mallia kehitettäessä on tärkeää huomioida minkälaisen absorptiospektrin esikäsittelyyn sitä rakennetaan. Mitä enemmän näytteestä ja mahdollisista virhelähteistä tiedetään, sitä parempi EMSC-malli voidaan rakentaa. Erityistä huomiota on syytä kiinnittää referenssispektrin valintaan sekä pääkomponenttien määrään.



## 8 Läheteet

- [1] A. Boskey and N. Pleshko Camacho, “FT-IR imaging of native and tissue-engineered bone and cartilage,” vol. 28, no. 15. pp. 2465–2478, 2007, doi: <https://doi-org.pc124152.oulu.fi:9443/10.1016/j.biomaterials.2006.11.043>.
- [2] M. V. Padalkar, R. G. Spencer, and N. Pleshko, “Near Infrared Spectroscopic Evaluation of Water in Hyaline Cartilage,” vol. 41, no. 11, pp. 2426–2436, 2013, doi: [10.1007/s10439-013-0844-0](https://doi-org.pc124152.oulu.fi:9443/10.1007/s10439-013-0844-0).
- [3] H. Martens, J. P. Nielsen, and S. Engelsen, “Light Scattering and Light Absorbance Separated by Extended Multiplicative Signal Correction. Application to Near-Infrared Transmission Analysis of Powder Mixtures,” vol. 75, no. 3, pp. 394–404, 2003, doi: [10.1021/ac020194w](https://doi-org.pc124152.oulu.fi:9443/10.1021/ac020194w).
- [4] P. Bassan *et al.*, “RMieS-EMSC correction for infrared spectra of biological cells: Extension using full Mie theory and GPU computing,” vol. 3, no. 8, pp. 609–620, 2010, doi: [10.1002/jbio.201000036](https://doi-org.pc124152.oulu.fi:9443/10.1002/jbio.201000036).
- [5] P. Bassan *et al.*, “FTIR microscopy of biological cells and tissue: data analysis using resonant Mie scattering (RMieS) EMSC algorithm,” vol. 137, no. 6, pp. 1370–1377, 2012, doi: [10.1039/C2AN16088A](https://doi-org.pc124152.oulu.fi:9443/10.1039/C2AN16088A).
- [6] A. Kohler *et al.*, “Estimating and Correcting Mie Scattering in Synchrotron-Based Microscopic Fourier Transform Infrared Spectra by Extended Multiplicative Signal Correction,” vol. 62, no. 3, pp. 259–266, 2008, doi: [10.1366/000370208783759669](https://doi-org.pc124152.oulu.fi:9443/10.1366/000370208783759669).
- [7] J. Skogholt, K. H. Liland, and U. G. Indahl, “Preprocessing of spectral data in the extended multiplicative signal correction framework using multiple reference spectra,” vol. 50, no. 3, pp. 407–417, 2019, doi: [10.1002/jrs.5520](https://doi-org.pc124152.oulu.fi:9443/10.1002/jrs.5520).
- [8] P. Bassan, H. J. Byrne, F. Bonnier, J. Lee, P. Dumas, and P. Gardner, “Resonant Mie scattering in infrared spectroscopy of biological materials – understanding the ‘dispersion artefact,’” vol. 134, no. 8, pp. 1586–1593, 2009, doi: [10.1039/B904808A](https://doi-org.pc124152.oulu.fi:9443/10.1039/B904808A).
- [9] T. Konevskikh, R. Lukacs, R. Blümel, A. Ponossov, and A. Kohler, “Mie scatter corrections in single cell infrared microspectroscopy,” vol. 187, no. 0, pp. 235–257, 2016, doi: [10.1039/C5FD00171D](https://doi-org.pc124152.oulu.fi:9443/10.1039/C5FD00171D).

- [10] J. H. Solheim, E. Gunko, D. Petersen, F. Großerüschkamp, K. Gerwert, and A. Kohler, “An open-source code for Mie extinction extended multiplicative signal correction for infrared microscopy spectra of cells and tissues,” vol. 12, no. 8, p. e201800415, 2019, doi: 10.1002/jbio.201800415.
- [11] N. K. Afseth and A. Kohler, “Extended multiplicative signal correction in vibrational spectroscopy, a tutorial,” vol. 117, pp. 92–99, 2012, doi: <https://doi-org.pc124152.oulu.fi:9443/10.1016/j.chemolab.2012.03.004>.
- [12] Å. Rinnan, F. van den Berg, and S. B. Engelsen, “Review of the most common pre-processing techniques for near-infrared spectra,” vol. 28, no. 10, pp. 1201–1222, 2009, doi: <https://doi.org/10.1016/j.trac.2009.07.007>.
- [13] T. Konevskikh, R. Lukacs, and A. Kohler, “An improved algorithm for fast resonant Mie scatter correction of infrared spectra of cells and tissues,” vol. 11, no. 1, p. e201600307, 2018, doi: 10.1002/jbio.201600307.
- [14] B. Zimmermann and A. Kohler, “Optimizing Savitzky–Golay Parameters for Improving Spectral Resolution and Quantification in Infrared Spectroscopy,” vol. 67, no. 8, pp. 892–902, 2013, doi: 10.1366/12-06723.
- [15] L. Rieppo, “Infrared Spectroscopic Characterization of Articular Cartilage,” University of Eastern Finland, 2012.
- [16] M. Féré *et al.*, “Focus on pre-processing step to ensure the clinical transferability of Raman data acquired on lymphocytes in different experimental and instrumental conditions,” vol. 103, p. 102931, 2019, doi: <https://doi.org/10.1016/j.vibspec.2019.102931>.
- [17] P. Bassan, *Gardnerlabuom/Rmies: Rmies\_Emsc\_V5\_Z*. Zenodo, 2017, doi: 10.5281/zenodo.832277